Development of a QSPR model for predicting thermal stabilities of nitroaromatic compounds taking into account their decomposition mechanisms

Guillaume Fayet, Patricia Rotureau, Laurent Joubert, Carlo Adamo

Development of QSPR models on thermal stability of nitroaromatic compounds considering their decomposition mechanisms

Guillaume Fayet[a,b*], Patricia Rotureau[a], Laurent Joubert[c] and Carlo Adamo[b]

[a] Institut National de l'Environnement Industriel et des Risques (INERIS), Parc Technologique Alata, BP2, 60550 Verneuil-en-Halatte, France

[b] Laboratoire d'Electrochimie, Chimie des Interfaces et Modélisation pour l'Energie, CNRS UMR-7575, Ecole nationale Supérieure de Chimie de Paris - Chimie ParisTech, 11 rue P. et M. Curie, 75231 Paris Cedex 05, France

[c] Institut de Recherche en Chimie Organique Fine (IRCOF), Université de Rouen, 76821 Mont Saint-Aignan, France

Abstract

The molecular structures of 77 nitroaromatic compounds have been correlated to their thermal stability by combining the Quantitative Structure-Property Relationship (QSPR) method with the Density Functional Theory (DFT). More than 300 descriptors (constitutional, topological, geometrical and quantum chemical) have been calculated and multilinear regressions have been performed to find accurate quantitative relationships with experimental heats of decomposition ($-\Delta H$). In particular, this paper demonstrated the importance of taking into account chemical mechanisms within the selection of an adequate experimental data set. A reliable QSPR model presenting strong correlation with experiments for both the training and the validation molecular sets (R^2=0.90 and 0.84, respectively) has been developed for non-ortho substituted nitroaromatic compounds. Moreover, its applicability domain was determined and the model's predictivity reached to 0.86 within this applicability domain. To our knowledge, this study led to the first QSPR model for the prediction of the thermal stability of energetic compounds, developed according to the OECD principles for regulatory acceptability.

Keywords: Quantitative Structure-Property Relationship (QSPR); nitroaromatic compounds; heat of decomposition; REACH regulation; Density Functional Theory (DFT)

*Corresponding author: guillaume.fayet@ineris.fr.

Introduction

After initiation from various external stimuli (impact, electric discharge, heat), energetic materials can undergo to decomposition reactions leading to the release of large amount of energy. So, the comprehension of their thermal stability properties is very important in order to evaluate not only their explosive power, but also the related hazards during storage, transportation and handling [1]. Indeed, many reported (industrial) accidents have been caused by the lack of knowledge about decomposition process [2], particularly those concerning nitro compounds [3].

To evaluate the amount of the energy released during the decomposition, calorimetric analyses are performed on chemicals. A typical experimental screening test is the differential scanning calorimetry (DSC) [1,4-6], which provides heats of decomposition with an error of about 5-10% [7].

Such measurement represents a pre-selection step for substances which may have explosive properties according to the Recommendations on the Transport of Dangerous Goods [8] which are used in REACH (for *"Registration, Evaluation, Authorisation and Restriction of Chemicals"*) [9] and CLP (for *"Classification, labelling and packaging of substances and mixtures"*) [10] regulations for classification purpose. Within this new European regulatory framework dedicated to chemicals, the evaluation of a tremendous number of substances may be required in a restricted calendar. Unlikely, the complete experimental characterization of chemicals causes not only time, cost and ethical (e.g. animals testing for toxicology and ecotoxicology) problems but also hazards for the case of potentially explosive compounds, (e.g. nitroaromatic compounds) [11]. In this context, the development of alternative tools, notably those issuing from computational chemistry [12], may be useful and is even recommended in these regulations for the screening and the prioritization of chemicals for experiments.

Among the available methods, Quantitative Structure-Property Relationships (QSPR) represent powerful tools to predict different properties of chemicals. Widely used in biology [13,14], toxicology [15,16] and drug design [17,18], their applications for physico-chemical properties increased since many years [19,20], particularly for the properties of energetic materials [21-32]. Their principle consists in developing a mathematic relationship connecting a macroscopic property of a compounds series to microscopic descriptors derived from their

molecular structures, using a reliable experimental data set. Once a model is developed and validated, it can be used to predict the macroscopic property for other molecules, with close structures, not yet characterized and even maybe not yet synthesized. Moreover, it might help to understand the investigated phenomena at a molecular scale. To encourage the development and use of QSAR/QSPR models, the Organization for Economic Co-operation and Development (OECD) introduced recently 5 principles for the validation, for regulatory purposes, of these models [33]. They require a defined endpoint, an unambiguous algorithm, a defined domain of applicability, appropriate measures of goodness-of-fit, robustness and predictivity and a mechanistic interpretation, if possible. Until now, only few QSPR models were investigated under these principles [34].

Some works were devoted to predict the thermal stability of chromophores [35], polymers [36] or ionic liquids [37]. Within the framework of energetic materials, some trends connecting the presence of particular functional groups to the decomposition temperatures (T_{onset}) of potentially explosive compounds were highlighted by Grewer in the early nineties [38]. The popular CHETAH software, dedicated to the prediction of reactivity hazards, estimates a maximum heat of decomposition of chemicals using a Benson's group contribution method [39]. Nevertheless, this quantity cannot be directly compared to experimental heats of decomposition since it considers a maximal decomposition of the molecule. More recently, T_{onset} was correlated to the dissociation energies of the weak bonds in nitro molecules [27,40].

To the best of our knowledge, the first QSPR type analysis related to the thermal stability of nitroaromatic compounds was realized by Saraf [26], based on 19 differential scanning calorimetric data. This study proposed the number of nitro groups (n_{NO2}) as the only variable to estimate the decomposition enthalpy ($-\Delta H$) with an average error about 8%, an incertitude close to the experimental error. More recently, Keshavarz proposed two QSPR models based on constitutional descriptors to predict the activation energy of thermolysis of nitroaromatics [24] and nitramines [23] with both R^2=0.87 in correlation.

In previous works, we proposed preliminary QSPR models to predict the heat of decomposition of 22 nitroaromatic compounds [41-43], the most robust one presenting good correlation with experimental data (R^2=0.98). Nevertheless, the lack of experimental data (only 22 molecules) did not allow the estimation of predictive power and applicability domains of the models.

This paper deals with the application of an original approach combining the QSPR methodology with the Density Functional Theory (DFT) for the prediction of the heat of decomposition of a larger series of nitroaromatic compounds (77 molecules), following the OECD recommendations for the validation of such predictive models. The chosen quantum chemical level of theory (DFT) allows for the optimization of molecular structures and, overall, the accurate calculations of a series of chemically-comprehensive descriptors of molecular reactivity (e.g. conceptual DFT descriptors) [41,43,42]. Furthermore, the importance of the training set selection is demonstrated and its definition will be in relation with the subjacent microscopic decomposition mechanisms.

Materials and methods

Experimental data set

The choice of the experimental data set is, of course, a critical point of any QSPR analysis. Since experimental conditions may have a strong influence, all experimental values used in the fitting procedure should be obtained in the same conditions to ensure, at best, their reliability and compatibility.

In this study, a data set of 77 nitroaromatic compounds was considered. It consists in nitrobenzenes derivatives including mono, di and trinitrobenzenes and a large variety of substituents (e.g. nitro, carboxylic acids or halogens). All the heats of decomposition (in tables 1 and 2) were extracted from a single reference [7] to ensure they were obtained using a single protocol. In the present case, a pressure DSC apparatus was used on 1-2 mg samples in aluminium cells with a pin-hole (i.e. open sample cells) with a heat rate of 10 K/min.

Molecular structures

All molecular structures were calculated using the density functional theory (DFT) with the Gaussian03 package [44]. Geometry optimizations were performed using the parameter-free PBE0 hybrid functional [45] and the 6-31+G(d,p) basis set. Vibrational frequencies were computed at the same theoretical level to ensure that all stable species corresponds to energy minima. These molecular structures were then loaded into CodessaPro software [46], including information about geometry, atomic charges, molecular orbitals energies and vibrational frequencies. More than 300 descriptors were calculated and, in addition, some particular external descriptors have been included since their interest has been shown in

previous work [41,42]. This is notably the case for the number of nitro groups and some conceptual DFT descriptors that have been already correlated successfully with thermal stability of similar nitroaromatic compounds [41,42].

The considered molecular descriptors can be divided into several classes: constitutional, topological, geometrical and quantum chemical descriptors. Constitutional descriptors characterize the presence and number of specific atoms, groups or bonds in the molecule (e.g. number of O atoms, single bonds). Topological indices, like Wiener index, are based on the atomic connectivity and give information about the size and branching degree of molecules. Geometric descriptors, contrary to the previous classes calculated from 2D-structures, stem from the 3D-structure (e.g. molecular volume). Quantum chemical descriptors gather binding, formation, molecular orbital energies, thermodynamic and electronic information, like dipole moment or polarizability, and descriptors characterizing the charge distribution within the molecule (e.g. partial charges). Detailed definitions and information about descriptors can be found in [47].

Statistical analysis

Various data mining tools can be used to develop QSPR models, e.g. artificial neural networks [48] or genetic algorithms [49]. In this study, multilinear regressions have been computed and the final model has the following general mathematic form:

$$Y = a_0 + \sum_{i=1}^{n} a_i X_i \qquad (1)$$

where Y is the property to predict, X_i are the molecular descriptors and a_i are the corresponding regression constants.

Equations (1) have been obtained using the “best multilinear regression” (BMLR) technique, described in [47] and implemented into CodessaPro software.

The first step of the BMLR analysis consists in reducing the initial set of descriptors by rejecting all descriptors with insignificant variance, to ensure that descriptors are not included by chance in the model whereas they are not related to the investigated property. At the same time, if two descriptors are highly correlated together, only the one presenting the best correlation with the property is kept. This step not only prevents against the introduction of

inappropriate descriptors, but also makes the analysis faster, since fewer variables have to be treated.

After that, starting from pairs of orthogonal (i.e. not intercorrelated) descriptors, higher rank models are computed by including orthogonal descriptors successively as soon as an increase in correlation is observed. To the end, the BMLR analysis selects the best models at each rank and the final model must be chosen between them. It has to be sufficiently correlated and, at the same time, ensure against any over-parameterization, which leads to a loss in predictive power for molecules outside the training set.

In this paper, the "breaking point" rule has been used to manage this problem. This method, already successfully used in previous works [22,50,43], consists in analyzing the correlation improvement with the number of variables in the model. By plotting the R^2 values as functions of the number of descriptors, an asymptotic behavior was observed and the improvement of correlation became less significant after a certain rank ($\Delta R^2 < 0.02$-0.03). At this point ("breaking point"), the model is considered as optimal, representing the best compromise between correlation and parameterization.

The robustness and the stability of the models have been evaluated through the square correlation coefficient (R^2) and the mean absolute error (MAE). The choice of the descriptors was confirmed by performing a student's t-test at a 95% confidence level. Moreover, they have been validated internally using the cross-validation technique (R^2_{cv}), and externally, using a validation set of data (R^2_{valid}).

Finally, the applicability domain of the models, i.e. the domain in which predictions are reliable, was investigated. In the present paper, the applicability domain estimation was performed based on Euclidean distances in the descriptors space (after a principal component analysis pre-processing) using Ambit Discovery software [51]. The applicability domain was defined to include 95% of the training set molecules and all predictions within this domain are expected to be reliable. This last statement was checked by calculating the correlation coefficients (R^2_{in}) for the molecules of the validation set inside this applicability domain.

Results and discussion

Complete data set

In a first step, the whole data set of 77 molecules has been investigated. This data set is sufficiently large to be divided into a training and a validation set. In order to keep similar distributions for the two ensembles, the entire set was sorted from the smallest to the highest values of heat of decomposition and the validation set was composed by the 3^{rd}, 7^{th}, etc. molecules. Thus, distributions of both training and validation sets (58 and 19 molecules, respectively) are quite similar to the entire one, as shown on figure 1. This ratio makes both sets sufficiently important for robust development and validation of the model. Then, the BMLR method was applied on the training set (see table 1). Models including up to 19 descriptors were built. The improvement in R^2 with the number of descriptors included in the model is represented in figure 2. Based on the "breaking point" rule, the following four-parameter model was considered as the best compromise between correlation and number of descriptors.

$$-\Delta H = -282.3 + 333.5\, n_{NO2} - 1214.5\, E_{C,avg} + 7.4\, \alpha - 275.6\, {}^{0}IC_{avg} \qquad (2)$$

where n_{NO2} is the number of nitro groups, $E_{C,avg}$ the average 1-electron reactivity index for a C atom, α the mean polarizability and ${}^{0}IC_{avg}$ the average information content (order 0), a topological index. This model presents the advantage to include descriptors directly related to thermal stability. In fact, the amount of energy released during decomposition is expected to be linked to the loss of nitro groups [52]. It is then pertinent to find the number of these nitro groups in the model. Moreover, $E_{C,avg}$ characterizes the reactivity of carbon atoms in the molecule and the one connected to leaving nitro groups is expected to be the most reactive in the molecule. Besides, nitro groups also influence the electronic properties of nitro groups, notably the polarizability.

This model is significantly correlated with R^2=0.84 and quite stable (R^2_{cv}=0.81). If it is less correlated than our previous work based on 22 molecules (R^2=0.98) [43], the distribution of experimental data was not as homogeneous as that considered in the present work and the single trinitro molecule of the previous set strongly influenced the regression.

To evaluate the predictive power of this new model, heats of decomposition of the validation set were calculated. As shown on figure 3 and table 2, the model gives only low correlation with an average deviation of 32% with experiments (R^2_{valid}=0.43). In particular, the calculated values for 4-fluoro-2-nitrotoluene, 2,4- and 2,6-dinitrobenzoic acids present large errors: 64, 83 and 206 %, respectively (see table 2). Nevertheless, no molecule of this validation set was

excluded from the applicability domain of the model (see table 3). So, the predictive power of the model is low within its own applicability domain (R^2_{in}=0.43).

At this point, standard experimental uncertainty is not sufficient to explain the poor predictive power of the model. Another factor affecting the accuracy of the model is related to the chemical reactivity. Indeed, the decomposition of nitroaromatic compounds is complex. Whereas the direct breaking of the carbon nitro bond was sometimes considered as the initiation step of decomposition for nitro compounds [40], numerous studies demonstrated that more complex reaction paths could be specifically involved in nitroaromatic molecules [52], such as ortho-substituted nitrobenzene derivatives (e.g. o-nitrotoluenes [53] and its derivatives [54] including 2,4,6-trinitrotoluene [55]).

For this reason, models have been developed considering non-ortho and ortho nitroaromatic compounds separately, i.e. a first model is applicable for compounds presenting no substituent in ortho position to the nitro group and a second one for compounds presenting one. So, two models have been developed following the protocol previously used for the entire set with same division into training and validation sets (in tables 1 and 2). Besides, Storm [56] and Kamlet [57] already considered the importance of the substituents in this position by distinguishing nitroaromatic explosives with an alpha C-H linkage when reporting experimental impact sensitivities and their correlation with the oxygen balance.

Non-ortho compounds

A four-parameter model was developed from the 31 non-ortho molecules of the training set:

$$-\Delta H = 0.8\ G - 3.8\ WPSA1 - 4255.1\ Q_{max} + 26.8\ RPCS - 251.2 \qquad (3)$$

where G is the gravitation index, WPSA1 the weighted positive surface area (from Mulliken calculated charges [58]), Q_{max} the maximal partial charge in the molecule (calculated according to Gasteiger's method [59]) and RPCS the relative positive charged surface area (from Zefirov calculated charges [47]). If Q_{max} is related to the nitrogen atoms in nitro groups, the other descriptors are more difficult to link to the decomposition process. G characterizes the molecular shape and the mass distribution within the molecule. WPSA1 and RPCS are charged partial surface area descriptors, developed to encode the features responsible for the polar interactions between molecules. Nevertheless, these last descriptors are related, in a certain way, to the distribution of charge within the molecule which is very influenced by

nitro groups. So, they are indirectly related to the properties of this group, which is central in the decomposition of nitroaromatic compounds.

This model is strongly correlated with experimental data (R^2=0.90) with an average deviation of 12%, close to the experimental uncertainty (see figure 4 and table 1). The internal validation, by cross-validation, is satisfying with R^2_{cv}=0.86. Moreover, it presents good predictivity since calculated values are close to experimental ones for the 11 non-ortho molecules of the validation set (R^2_{valid}=0.84, see table 2). Regarding to experimental uncertainty, the performance of this model is satisfying since the predictive power of the model reaches even to 0.86 when excluding the molecule determined to be out of the applicability domain of the model (see table 3).

Ortho compounds

For ortho nitroaromatic compounds, the final model, developed on the 27 molecules of the training set, is another four-parameter equation:

$$-\Delta H = 4.1\ PNSA1 - 3298.8\ RPCG - 56228\ N_{C,min} - 1245.5\ S_{ZX/ZX} + 1117.8 \qquad (4)$$

where PNSA1 and RPCG are the partial negative charged surface area and the relative positive charge (from Zefirov calculated charges), $N_{C,min}$ is the minimum nucleophilic reactivity index for a C atom and $S_{ZX/ZX}$ (ZX shadow / ZX rectangle) characterizes the molecular shape within a ZX plane. The charge related descriptors are the main descriptors in this equation. The only descriptor being directly related to the thermal stability is $N_{C,min}$, which is influenced by the electronic properties of nitro groups, so to their reactivity with the aromatic ring.

Concerning the performance of the model, the correlation with experimental data is high (R^2=0.94) with an average deviation of 17% and the cross-validation procedure exhibits a good robustness (R^2_{cv}=0.91) but the predictivity is low since calculated values for the 8 ortho molecules of the validation set deviate about 44% from experiment (R^2_{valid}=0.42).

In fact, a reliable model was more expected for the non-ortho than for the ortho molecules since many different reaction paths exist upon the nature of the substituent in ortho position from the nitro group (nitro, alcohol, amino…) [52] whereas all non-ortho nitroaromatic compounds follow the same decomposition process, C-NO_2 homolysis being evidenced in previous theoretical studies [60,61] as the main reaction path (without influence of the

substituent nature in meta and para position to the nitro group). Besides the applicability domain analysis confirms this observation since the predictive power (R^2_{in}) of models 2 (entire set) and 4 (ortho molecules) are low in their applicability domain, compared to the one of model 3 (non-ortho molecules).

To compare the models developed in this study to previous ones, the heats of decomposition of the validation molecules have been calculated using Saraf's model [26] (based on the only number of nitro groups of 19 nitrobenzene derivatives) and to our previous model [43] (based on 22 molecules), as shown in table 2. No satisfactory correlation was exhibited since the average deviations are 36% and 46% (R^2= 0.53 and 0.37) for Saraf's model and our previous work, respectively. This may be due to the lack of homogeneity in the data sets, which moreover consider no or only one trinitro compounds. Besides, these models did not distinguished ortho and non-ortho compounds. Finally, the predictive powers of these models are very low, even in their respective applicability domains (as shown in table 3). These previous models present the same limit as Eq. 2 and 4 by considering different decomposition mechanisms in the same data set. So, the knowledge of molecular decomposition paths is a critical point for the prediction of nitroaromatics thermal stabilities.

Accordance with OECD principles

Our best QSPR model, developed for non-ortho compounds (Eq. 3) follows the five OECD principles of validation of QSAR/QSPR models for regulatory use [33]:

Principle 1: The endpoint is well defined as the heat of decomposition, performed using a pressure DSC apparatus with a 10 K/min heat rate on 1-2mg samples in aluminum cells with pin-hole [7].

Principle 2: The model is transparent since it consists in a simple multi-linear equation, including four parameters, calculated from well defined DFT-optimized structures at the PBE0/6-31+G(d,p) level.

Principle 3: The model is applicable to all nitrobenzene derivatives, without substituent in ortho position to the nitro group, that are included into an applicability domain defined by interpolation of the training set

Principle 4: The performance of the model was estimated by its correlation in the training set (R^2=0.90), its robustness by cross validation (R^2_{cv}=0.86) and its predictive power into its domain of applicability on an external validation set of molecules (R^2_{in}=0.86).

Principle 5: If no mechanistic interpretation can be easily provided from descriptors used in the model, the molecular mechanisms were considered during its development and form part of the defined applicability domain, since the model is not applicable to ortho-substituted compounds (see Principle 3). Indeed, the molecular mechanism involved in the decomposition of the target non-ortho substituted nitroaromatic compounds has been characterized from density functional theory calculations in previous work [57].

As this model satisfies all requirements of OECD principles, it could be used as an efficient alternative to experimental characterization as a first screening test to evaluate if a target nitroaromatic compound may have explosive properties.

Conclusion

A set of 77 nitroaromatic compounds was considered aiming to develop a reliable QSPR model for the prediction of their heats of decomposition. Molecular structures, calculated at a DFT level of theory, were described using more than 300 descriptors.

Considering the entire set of data led to significant correlation in the training step (R^2=0.84) but it failed in predicting the heats of decomposition of external molecules (about 32% in deviation). If experimental uncertainty (e.g. use of open sample cells) contributed to the lack of predictive power, this paper demonstrated that the data set selection had to take into account chemical reactivity during the decomposition process.

Indeed, considering a set of nitro compounds without any ortho substituent to the nitro group led to a performant model with significant correlation between calculated and experimental heats of decomposition of molecules out of the training set (R^2_{valid}=0.84). Besides, within its own domain of applicability, the predictivity of this model is very high (R^2_{in}=0.86). The improvement of accuracy, compared with the model for the entire set, is related to the fact that various specific decomposition mechanisms involve when substituents are in ortho position from the nitro groups whereas all non-ortho compounds decomposed following the same reaction path (homolysis of the C-NO_2 bond).

Finally, this model is, to our knowledge, the most reliable QSPR model predicting the heat of decomposition within the class of nitroaromatic compounds and the first dedicated to energetic compounds following all OECD requirements for regulatory use.

References

[1] Grewer T (1994) Thermal Hazards of Chemical Reactions. Elsevier, Amsterdam

[2] Etchells JC (1997) Why Reactions Run Away. Org Process Res Dev 1: 435-437.

[3] Gustin JL (1998) Runaway Reaction Hazards in Processing Organic Nitro Compounds. Org Process Res Dev 2: 27-33.

[4] Jones DEG, Augsten RA (1996) Evaluation of systems for use in DSC measurements on energetic materials. Thermochim Acta 286: 355-373.

[5] Chervin S, Bodman GI (2003) Method for estimating decomposition characteristics of energetic chemicals. Process Saf Prog 22: 241-243.

[6] Yoshida T (1987) Safety of Reactive Chemicals, vol 1. Industrial Safety Series. Elsevier, Amsterdam

[7] Ando T, Fujimoto Y, Morisaki S (1991) Analysis of differential scanning calorimetric data for reactive chemicals. J Hazard Mater 28: 251-280.

[8] United Nations (2009) Recommendations on the Transport of Dangerous Goods: Manual of Tests and Criteria, Fifth revised edition, ST/SG/AC.10/11/Rev.5, Geneva/New York.

[9] Regulation (EC) N° 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH).

[10] Regulation (EC) N°1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) N° 1907/2006.

[11] Medard L (1987) Les explosifs occasionnels. Tech. & Doc., Cachan, France

[12] Lewis A, Kazantzis N, Fishtik I, Wilcox J (2007) Integrating process safety with molecular modeling-based risk assessment of chemicals within the REACH regulatory framework: Benefits and future challenges. J Hazard Mater 142: 592-602.

[13] Gao H, Katzenellenbogen JA, Garg R, Hansch C (1999) Comparative QSAR Analysis of Estrogen Receptor Ligands. Chem Rev 99: 723-744.

[14] Winkler DA (2002) The role of quantitative structure - activity relationships (QSAR) in biomolecular discovery. Brief Bioinf 3: 73-86.

[15] Bradbury SP (1995) Quantitative structure-activity relationships and ecological risk assessment: an overview of predictive aquatic toxicology research. Toxicol Lett 79: 229-237.

[16] Selassie CD, Garg R, Kapur S, Kurup A, Verma RP, Mekapati SB, Hansch C (2002) Comparative QSAR and the Radical Toxicity of Various Functional Groups. Chem Rev 102: 2585-2606.

[17] Grover M, Singh B, Bakshi M, Singh S (2000) Quantitative structure-property relationships in pharmaceutical research - Part 2. Pharm Sci Tech Today 3: 50-57.

[18] Grover M, Singh B, Bakshi M, Singh S (2000) Quantitative structure-property relationships in pharmaceutical research - Part 1. Pharm Sci Tech Today 3: 28-35.

[19] Cocchi M, De Benedetti PG, Seeber R, Tassi L, Ulrici A (1999) Development of Quantitative Structure-Property Relationships Using Calculated Descriptors for the Prediction of the Physicochemical Properties (n_D, ρ, bp, ε, η) of a Series of Organic Solvents. J Chem Inf Comput Sci 39: 1190-1203.

[20] Katritzky AR, Lobanov VS, Karelson M (1995) QSPR: the Correlation and Quantitative Prediction of Chemical and Physical-Properties from Structure. Chem Soc Rev 24: 279-287.

[21] Katritzky AR, Petrukhin R, Jain R, Karelson M (2001) QSPR Analysis of Flash Points. J Chem Inf Comput Sci 41: 1521-1530.

[22] Katritzky AR, Stoyanova-Slavova IB, Dobchev DA, Karelson M (2007) QSPR modeling of flash points: An update. J Mol Graph Model 26: 529-536.

[23] Keshavarz MH (2009) Simple method for prediction of activation energies of the thermal decomposition of nitramines. J Hazard Mater 162: 1557-1562.

[24] Keshavarz MH, Pouretedal HR, Shokrolahi A, Zali A, Semnani A (2008) Predicting activation energy of thermolysis of polynitro arenes through molecular structure. J Hazard Mater 160: 142-147.

[25] Rice BM, Hare JJ (2002) A Quantum Mechanical Investigation of the Relation between Impact Sensitivity and the Charge Distribution in Energetic Molecules. J Phys Chem A 106: 1770-1783.

[26] Saraf SR, Rogers WJ, Mannan MS (2003) Prediction of reactive hazards based on molecular structure. J Hazard Mater 98: 15-29.

[27] Theerlynck E, Mathieu D, Simonetti P (2005) Towards improved models to rationalize and estimate the decomposition temperatures of nitroalkanes, nitramines and nitric esters. Thermochim Acta 426: 123-129.

[28] Afanas'ev GT, Pivina TS, Sukhachev DV (1993) Comparative characteristics of some experimental and computational methods for estimating Impact Sensitivity Parameters of Explosives. Propel Explos Pyrotech, 18: 309-316.

[29] Toghiani RK, Toghiani H, Maloney SW, Boddu VM (2008) Prediction of physicochemical properties of energetic materials. Fluid Phase Eq 264: 86-92.

[30] Politzer P, Ma Y, Lane P, Concha MC (2005) Computational Prediction of Standard Gas, Liquid, and Solid-Phase Heats of Formation and Heats of Vaporization and Sublimation. Int J Quantum Chem 106: 341-347.

[31] Politzer P, Martinez J, Murray JS, Concha MC, Toro-Labbe A (2009) An electrostatic interaction correction for improved crystal density prediction. Mol Phys 107: 2095-2101.

[32] Pospisil M, Vavra P, Concha MC, Murray JS, Politzer P (2010) A possible crystal volume factor in the impact sensitivities of some energetic compounds. J Mol Model 16: 895-901.

[33] Principles for the Validation, for Regulatory Purposes, of (Quantitative) Structure-Activity Relationship Models (2009). Organisation for Economic Co-operation and Development (OECD), Paris

[34] Papa E, Kovarich S, Gramatica P (2009) Development, Validation and Inspection of the Applicability Domain of QSPR Models for Physicochemical Properties of Polybrominated Diphenyl Ethers. QSAR Comb Sci 28: 790-796.

[35] Figueiredo LJO, Garrido FMS (2001) Chemometric analysis of nonlinear optical chromophores structure and thermal stability. J Mol Struct (THEOCHEM) 539: 75-81.

[36] Yu X, Xie Z, Yi B, Wang X, Liu F (2007) Prediction of the thermal decomposition property of polymers using quantum chemical descriptors. Eur Polym J 43: 818-823.

[37] Kroon MC, Buijs W, Peters CJ, Witkamp G-J (2007) Quantum chemical aided prediction of the thermal decomposition mechanisms and temperatures of ionic liquids. Thermochim Acta 465: 40-47.

[38] Grewer T (1991) The influence of chemical structure on exothermic decomposition. Thermochim Acta 187: 133-149.

[39] Benson SW (1976) Thermochemical kinetics. 2nd edn. Wiley, New York

[40] Saraf SR, Rogers WJ, Mannan MS (2003) Application of Transition State Theory for Thermal Stability Prediction. Ind Eng Chem Res 42: 1341-1346.

[41] Fayet G, Joubert L, Rotureau P, Adamo C (2009) On the use of descriptors arising from the conceptual density functional theory for the prediction of chemicals explosibility. Chem Phys Lett 467: 407-411.

[42] Fayet G, Rotureau P, Joubert L, Adamo C (2009) On the prediction of thermal stability of nitroaromatic compounds using quantum chemical calculations. J Hazard Mater 171: 845-850.

[43] Fayet G, Rotureau P, Joubert L, Adamo C (2010) QSPR Modeling of Thermal Stability of Nitroaromatic Compounds: DFT vs. AM1 Calculated Descriptors. J Mol Model 16: 805-812.

[44] Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA, Jr., Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2004) Gaussian03. Gaussian Inc., Wallington CT

[45] Adamo C, Barone V (1999) Toward reliable density functional methods without adjustable parameters: The PBE0 model. J Chem Phys 110: 6158-6170.

[46] CodessaPro (2002). University of Florida, www.codessa-pro.com.

[47] Karelson M (2000) Molecular Descriptors in QSAR/QSPR. Wiley, New York

[48] Gasteiger J, Zupan J (1993) Neural Networks in Chemistry. Angew Chem Int Ed Engl 32: 503-527.

[49] Leardi R (2001) Genetic algorithms in chemometrics and chemistry: a review. J Chemometr 15: 559-569.

[50] Katritzky AR, Dobchev DA, Tulp I, Karelson M, Carlson DA (2006) QSAR study of mosquito repellents using Codessa Pro. Bioorg Med Chem Lett 16: 2306-2311.

[51] Jaworska J, Jeliazkova N (2007) Ambit Discovery 1.20.

[52] Brill TB, James KJ (1993) Kinetics and mechanisms of thermal decomposition of nitroaromatic explosives. Chem Rev 93: 2667-2692.

[53] Chen SC, Xu SC, Diau E, Lin MC (2006) A Computational Study on the Kinetics and Mechanism for the Unimolecular Decomposition of o-Nitrotoluene. J Phys Chem A 110: 10130-10134.

[54] Fayet G, Joubert L, Rotureau P, Adamo C (2009) A theoretical study of the decomposition mechanisms on substituted ortho-nitrotoluenes. J Phys Chem A 113: 13621-13627.

[55] Cohen R, Zeiri Y, Wurzberg E, Kosloff R (2007) Mechanism of Thermal Unimolecular Decomposition of TNT (2,4,6-Trinitrotoluene): A DFT Study. J Phys Chem A 111: 11074-11083.

[56] Storm CP, Stine JR (1990) Sensitivity Relationships in Energetic Materials. In: Bulusu SN (ed) Chemistry and Physics of Energetic Materials. Kluwer Academic Publishers, Netherlands, pp 605-630.

[57] Kamlet MJ, Adolph HG (1979) The relationship of Impact Sensitivity with Structure of Organic High Explosives. II. Polynitroaromatic explosives. Propel Explos 4: 30-34.

[58] Mulliken RS (1955) Electronic Population Analysis on LCAO-MO Molecular Wave Functions. I. J Chem Phys 23: 1833-1840.

[59] Gasteiger J, Marsili M (1980) Iterative partial equalization of orbital electronegativity--a rapid access to atomic charges. Tetrahedron 36: 3219-3228.

[60] Brill TB, James KJ, Chawla R, Nicol G, Shukla A, Futrell JH (1999) Influence of the substituent on the major decomposition channels of the NO2 group in para-substituted nitrobenzenes: a tandem mass spectrometric study. J Phys Org Chem 12: 819-826.

[61] Fayet G, Joubert L, Rotureau P, Adamo C (2008) Theoretical Study of the Decomposition Reactions in Substituted Nitrobenzenes. J Phys Chem A 112: 4054-4059.

Table 1. Experimental and calculated heats of decomposition (-ΔH) in kJ/mol for the molecules of the training set

molecules	Exp [7]	Eq. 2	Eq. 3	Eq. 4
non-ortho compounds				
2-amino-4-nitrophenol	130	238	173	-
3-nitrotoluene	149	238	212	-
2-amino-5-nitrophenol	153	201	239	-
4-nitrotoluene	213	253	192	-
4-nitrophenol	232	183	235	-
3-nitroanisole	243	223	288	-
3-nitrobenzoic acid methyl ester	256	334	277	-
2,6-dichloro-4-nitroaniline	264	394	284	-
4-nitrophenetole	270	347	249	-
4-nitrophenylhydrazine	277	389	279	-
3-nitrophenol	283	165	227	-
3-nitrobenzoic acid	289	267	372	-
4-nitroacetophenone	291	348	343	-
4-nitrobenzyl alcohol	292	223	272	-
4-nitrobenzoic acid methyl ester	302	329	264	-
4-nitro-2-toluidine	306	287	315	-
4-nitrobenzamide	319	275	321	-
4-nitrobenzyl chloride	337	675	333	-
4-nitroaniline	347	298	308	-
3-nitroaniline	350	302	317	-
3-nitrophenylacetic acid	358	397	347	-
4-nitrobenzhydrazide	362	415	335	-
3-nitroacetoanilide	369	289	394	-
2-amino-4-nitroanisole	375	339	325	-
4-nitroacetoanilide	387	343	372	-
4-nitrobenzoyl chloride	408	463	303	-
3-nitrocinnamic acid	414	314	417	-
4-nitrobenzaldehyde	421	495	394	-
3,5-dinitrobenzonitrile	654	699	698	-
3,5-dinitrobenzoic acid	674	658	679	-
3,5-dinitrobenzylchloride	711	682	673	-
ortho compounds				
5-chloro-2-nitrobenzotrifluoride	40	96	-	7
2-nitrophenol	123	176	-	172
2-nitrophenylacetic acid	175	288	-	266
2-nitrotoluene	182	223	-	211
2-nitroanisole	230	203	-	276
2-nitrobenzamide	256	317	-	341
2-nitrobenzoic acid	271	212	-	247
2-nitrobenzoic acid methyl ester	274	284	-	285

2-nitroaniline	280	267	-	284
2-nitroacetophenone	308	243	-	244
2-nitrobenzaldehyde	318	196	-	351
2-nitrobenzyl alcohol	319	350	-	239
1-chloro-3,4-dinitrobenzene	342	358	-	331
2-nitrophenylsulfenyl chloride	344	302	-	276
4-chloro-2-nitroaniline	349	263	-	312
2-nitrophenylhydrazine	381	448	-	298
1,5-dinitro-2,4-difluorobenzene	439	396	-	483
5-nitrovanillin	450	448	-	491
2-nitro-5-thiocyanotobenzoic acid	492	718	-	559
2,4-dinitroaniline	597	520	-	694
2-nitrobenzhydrazide	598	550	-	463
2-nitrocinnamic acid	600	620	-	606
3,4-dinitrotoluene	684	645	-	626
3,4-dinitrobenzoic acid	701	706	-	803
2,6-dinitroaniline	719	713	-	619
2-chloro-3,5-dinitrobenzoic acid	1023	766	-	937
2,4,6-trinitrotoluene	1223	1149	-	1220
	R^2	0.84	0.90	0.94
	MAE (%)	23	12	17
	R^2_{cv}	0.81	0.86	0.91

Table 2. Experimental and calculated heats of decomposition (-ΔH) in kJ/mol for the molecules of the validation set

molecules	Exp [7]	Eq. 2	Eq. 3	Eq. 4	Saraf [26]	Previous works [43]
			non-ortho compounds			
nitrobenzene	161	188	202	-	314	319
4-nitroanisole	248	248	283	-	314	312
4-nitrophenylacetic acid	265	291	341	-	314	314
3-nitroacetophenone	276	338	364	-	314	308
4-nitrobenzoic acid	284	275	332	-	314	267
3-nitrobenzamide	311	298	334	-	314	39
3-nitrobenzyl alcohol	325	259	258	-	314	352
3-nitrobenzaldehyde	373	297	389	-	314	339
3-nitrobenzhydrazide	430	722	344	-	627	631
4-nitrocinnamic acid	506	355	414	-	314	413
3,5-dinitrobenzamide	736	721	687	-	627	560
			ortho compounds			
4-fluoro-2-nitrotoluene	129	212	-	211	314	354
2,6-dinitrobenzoic acid	222	681	-	423	627	607
2-nitroacetoanilide	297	431	-	472	314	80
4-nitro-3-cresol	345	204	-	203	314	319
4-chloro-3-nitrobenzoic acid	354	238	-	463	314	325
2,4-dinitrobenzoic acid	394	306	-	553	314	280
2,4-dinitrotoluene	632	585	-	607	314	205
3,4-dinitrobenzylalcohol	685	660	-	683	627	605
	R^2_{valid}	0.43	0.84	0.42	0.53	0.37
	MAE (%)	32	18	44	36	46

Table 3. Analysis of the applicability domain of QSPR models

	Training set					Validation set				
	R^2	N_{in}	N_{out}	$\%_{in}$	R^2_{in}	R^2_{valid}	N_{in}	N_{out}	$\%_{in}$	R^2_{in}
Eq.2	0.84	55	3	95	0.82	0.43	19	0	100	0.43
Eq.3	0.90	29	2	94	0.88	0.84	10	1	90	0.86
Eq.4	0.94	26	1	96	0.94	0.42	8	0	100	0.42
Saraf [26]	0.98	19	0	100	0.98	0.53	17	2	89	0.54
Previous works [43]	0.98	21	1	95	0.99	0.37	15	4	79	0.24

N_{in}= count of molecules into the applicability domain

N_{out}= count of molecules out of the applicability domain

$\%_{in}$= ratio of molecules into the applicability domain

R^2_{in}= predictive power into the applicability domain

Figure captions

Figure 1. Distributions of the experimental heat of decomposition in the entire, training and validation sets

Figure 2. Number of descriptors versus R^2 of the models from the BMLR analysis on the entire training set.

Figure 3. Experimental versus calculated heats of decomposition (in kJ/mol) of nitroaromatic compounds according to Eq. 2 (the molecules of the training and validation sets are represented in wide triangles and plain circles respectively).

Figure 4. Experimental versus calculated heats of decomposition (in kJ/mol) of non-ortho nitroaromatic compounds according to Eq. 3 (the molecules of the training and validation sets are represented in wide triangles and plain circles respectively).